

FreeBayes Variants on Yeast

Cesky Krumlov

January 8, 2020

Running FreeBayes for Variant Calling

We can run FreeBayes to call variants on our yeast Illumina data (you could probably also run this on the error corrected PacBio data, but I'm not sure what would happen if you ran on the uncorrected data). To do this:

```
freebayes -f S288C.fa <bamfile> > <vcffile.vcf>
```

Let's take a look at the vcf.

A lot of the variants have genotype 0/0. That is, they're homozygous reference. FreeBayes provides no default filtering. We can get rid of most of these because they're low quality. We will use a tool from the package vcflib called vcffilter:

```
vcffilter -f 'QUAL > 10' <vcffile.vcf> > <filtered.vcf>
```

After Filtering

After filtering, you'll notice that all the 0/0 calls are gone, but now some of the calls are 0/1, which we wouldn't expect because we know that this yeast sample is essentially haploid. By default, FreeBayes assumes samples are diploid, but it has a nice ploidy option, so if we rerun with that, we should get something a little more reasonable:

```
freebayes -f S288C.fa -ploidy 1 <bamfile> > <haploid.vcf>
```

```
vcffilter -f 'QUAL > 10' <haploid.vcf> > <haploidfiltered.vcf>
```

Load Into IGV

You can load these vcf files into IGV and see what the variants look like compared to the read alignments. We will do more of this in the human example, but it might be interesting to see some of them here. The variants are likely to be in the regions I've noted as interesting, but some might be in other regions. You can load both the haploid and diploid calls and compare them.